



Abbott, S. (2019). getTBinR: an R package for accessing and summarising the World Health Organisation Tuberculosis data. *Journal of Open Source Software*, 4(34), [1260].
<https://doi.org/10.21105/joss.01260>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.21105/joss.01260](https://doi.org/10.21105/joss.01260)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Open Journals at <https://joss.theoj.org/papers/10.21105/joss.01260>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

getTBinR: an R package for accessing and summarising the World Health Organisation Tuberculosis data

Sam Abbott¹

¹ Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, UK

DOI: [10.21105/joss.01260](https://doi.org/10.21105/joss.01260)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 29 January 2019

Published: 26 February 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

Developing tools for rapidly accessing and exploring data sets benefits the public health research community by enabling multiple data sets to be combined in a consistent manner, increasing the visibility of key data sets, and providing a framework that can be used to explore key questions of interest. Tooling also reduces the barriers to entry, allowing non-specialists to explore data sets that would otherwise be inaccessible. This widening of access may also lead to new insights and wider interest for key public health issues.

`getTBinR` is an R package (R Core Team, 2019) to facilitate working with the data (World Health Organisation, 2018) collected by the World Health Organisation (WHO) on the country level epidemiology of Tuberculosis (TB). All data is freely available from the [WHO](#) and the package code is archived on Zenodo (Abbott, 2019) and [Github](#). The aim of `getTBinR` is to allow researchers, and other interested individuals, to quickly and easily gain access to a detailed TB data set and to start using it to derive key insights. It provides a consistent set of tools that can be used to rapidly evaluate hypotheses on a widely used data set before they are explored further using more complex methods or more detailed data. The functions provided in this package were developed to have sensible defaults to allow those new to the field to quickly gain key insights but also allow sufficient customisation so that experienced users may rapidly prototype new ideas.

The data sourced by `getTBinR` is collected by the WHO, via member governments, and used to compile the yearly global TB report (World Health Organisation, 2018). Data collection encompasses TB incidence, TB mortality rates, the age distribution of TB cases, the proportion of drug resistant cases, case detection rates, and treatment rates. For a complete description of the data and data collection process, see (World Health Organisation, 2018). These data are used by the WHO, governmental organisations and researchers to summarise country level TB epidemiology, as well as the wider global and regional picture.

The `getTBinR` package facilitates downloading the most up-to-date version of multiple TB relevant data sources from the WHO, along with the accompanying data dictionaries. It also contains functions to allow easy exploration of the data via searching data dictionaries, summarising key metrics on a regional and global level, and visualising the data in a variety of highly customisable ways. In addition, it provides both a dashboard and an automated country level report that enables the global, regional, and country level picture to be quickly summarised. An example of a potential use of the package is to explore estimates of the TB case fatality ratio (Abbott, 2018). In a few lines of code, using only built in package tooling, large regional differences can be discovered. Further insights can then be found by linking to other publicly available data sets or using a model based analysis. See <https://www.samabbott.co.uk/getTBinR/> for documentation.

Acknowledgements

Thanks to the staff at the World Health Organisation for compiling the data sets used in this package. SA is funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Evaluation of Interventions at University of Bristol in partnership with Public Health England (PHE). The views expressed are those of the author and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

References

- Abbott, S. (2018, February). Exploring estimates of the tuberculosis case fatality ratio - with getTBinR. Retrieved from https://www.samabbott.co.uk/getTBinR/articles/case_fatality_rate.html
- Abbott, S. (2019, January). Seabbs/getTBinR 0.5.8. *Zenodo*. doi:[10.5281/zenodo.2547405](https://doi.org/10.5281/zenodo.2547405)
- R Core Team. (2019). R: A language and environment for statistical computing. Retrieved from <https://www.R-project.org/>
- World Health Organisation. (2018, October). Global tuberculosis report. Retrieved from <https://www.who.int/tb/country/data/download/en/>